

SEQUENCE LISTING



(1) GENERAL INFORMATION:

(i) APPLICANT: CAPUT, DANIEL

FERRARA, PASCUAL

GUILLEMOT, JEAN-CLAUDE

KAGHAD, MOURAD

LEGOUX, RICHARD

LOISON, GERARD

LARBRE, ELIZABETH

LUPKER, JOHANNES

LEPLATOIS, PASCUAL

SALOME, MARK

(ii) TITLE OF INVENTION: URATE OXIDASE ACTIVITY PROTEIN,
RECOMBINANT GENE CODING THEREFOR, EXPRESSION VECTOR,
MICRO-ORGANISMS AND TRANSFORMED CELLS

(iii) NUMBER OF SEQUENCES: 36

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Foley & Lardner

(B) STREET: 1800 Diagonal Road, Suite 500

(C) CITY: Alexandria

(D) STATE: Virginia

(E) COUNTRY: USA

(F) ZIP: 22313-0299

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 07/659,408

(B) FILING DATE: 25-APR-1991

(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: BENT, Stephen A.

(B) REGISTRATION NUMBER: 29,768

(C) REFERENCE/DOCKET NUMBER: 16781/276 BEDL

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (703)836-9300

(B) TELEFAX: (703)683-4109

(C) TELEX: 899149

(2) INFORMATION FOR SEQ ID NO:1:

Patent Application US/07/659,408A

54
55 (i) SEQUENCE CHARACTERISTICS:
56 (A) LENGTH: 301 amino acids
57 (B) TYPE: amino acid
58 (D) TOPOLOGY: linear
59
60 (ii) MOLECULE TYPE: protein
61
62 (iii) HYPOTHETICAL: NO
63
64 (vi) ORIGINAL SOURCE:
65 (A) ORGANISM: Aspergillus flavus
66
67 (vii) IMMEDIATE SOURCE:
68 (B) CLONE: Urate oxidase
69
70
71 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
72
73 Ser Ala Val Lys Ala Ala Arg Tyr Gly Lys Asp Asn Val Arg Val Tyr
74 1 5 10 15
75
76 Lys Val His Lys Asp Glu Lys Thr Gly Val Gln Thr Val Tyr Glu Met
77 20 25 30
78
79 Thr Val Cys Val Leu Leu Glu Gly Glu Ile Glu Thr Ser Tyr Thr Lys
80 35 40 45
81
82 Ala Asp Asn Ser Val Ile Val Ala Thr Asp Ser Ile Lys Asn Thr Ile
83 50 55 60
84
85 Tyr Ile Thr Ala Lys Gln Asn Pro Val Thr Pro Pro Glu Leu Phe Gly
86 65 70 75 80
87
88 Ser Ile Leu Gly Thr His Phe Ile Glu Lys Tyr Asn His Ile His Ala
89 85 90 95
90
91 Ala His Val Asn Ile Val Cys His Arg Trp Thr Arg Met Asp Ile Asp
92 100 105 110
93
94 Gly Lys Pro His Pro His Ser Phe Ile Arg Asp Ser Glu Glu Lys Arg
95 115 120 125
96
97 Asn Val Gln Val Asp Val Val Glu Gly Lys Gly Ile Asp Ile Lys Ser
98 130 135 140
99
100 Ser Leu Ser Gly Leu Thr Val Leu Lys Ser Thr Asn Ser Gln Phe Trp
101 145 150 155 160
102
103 Gly Phe Leu Arg Asp Glu Tyr Thr Thr Leu Lys Glu Thr Trp Asp Arg
104 165 170 175
105
106 Ile Leu Ser Thr Asp Val Asp Ala Thr Trp Gln Trp Lys Asn Phe Ser

Patent Application US/07/659,408A

	180	185	190
107			
108			
109	Gly Leu Gln Glu Val Arg Ser His Val Pro Lys Phe Asp Ala Thr Trp		
110	195	200	205
111			
112	Ala Thr Ala Arg Glu Val Thr Leu Lys Thr Phe Ala Glu Asp Asn Ser		
113	210	215	220
114			
115	Ala Ser Val Gln Ala Thr Met Tyr Lys Met Ala Glu Gln Ile Leu Ala		
116	225	230	235
117			
118	Arg Gln Gln Leu Ile Glu Thr Val Glu Tyr Ser Leu Pro Asn Lys His		
119	245	250	255
120			
121	Tyr Phe Glu Ile Asp Leu Ser Trp His Lys Gly Leu Gln Asn Thr Gly		
122	260	265	270
123			
124	Lys Asn Ala Glu Val Phe Ala Pro Gln Ser Asp Pro Asn Gly Leu Ile		
125	275	280	285
126			
127	Lys Cys Thr Val Gly Arg Ser Ser Leu Lys Ser Lys Leu		
128	290	295	300
129			

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 302 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Aspergillus flavus*

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Met-Urate oxidase

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ser Ala Val Lys Ala Ala Arg Tyr Gly Lys Asp Asn Val Arg Val			
1	5	10	15
Tyr Lys Val His Lys Asp Glu Lys Thr Gly Val Gln Thr Val Tyr Glu			
20	25	30	
Met Thr Val Cys Val Leu Leu Glu Gly Glu Ile Glu Thr Ser Tyr Thr			
35	40	45	

Patent Application US/07/659,408A

```
160 Lys Ala Asp Asn Ser Val Ile Val Ala Thr Asp Ser Ile Lys Asn Thr
161      50                      55                      60
162
163 Ile Tyr Ile Thr Ala Lys Gln Asn Pro Val Thr Pro Pro Glu Leu Phe
164      65                      70                      75                      80
165
166 Gly Ser Ile Leu Gly Thr His Phe Ile Glu Lys Tyr Asn His Ile His
167                      85                      90                      95
168
169 Ala Ala His Val Asn Ile Val Cys His Arg Trp Thr Arg Met Asp Ile
170                      100                      105                      110
171
172 Asp Gly Lys Pro His Pro His Ser Phe Ile Arg Asp Ser Glu Glu Lys
173                      115                      120                      125
174
175 Arg Asn Val Gln Val Asp Val Val Glu Gly Lys Gly Ile Asp Ile Lys
176      130                      135                      140
177
178 Ser Ser Leu Ser Gly Leu Thr Val Leu Lys Ser Thr Asn Ser Gln Phe
179      145                      150                      155                      160
180
181 Trp Gly Phe Leu Arg Asp Glu Tyr Thr Thr Leu Lys Glu Thr Trp Asp
182                      165                      170                      175
183
184 Arg Ile Leu Ser Thr Asp Val Asp Ala Thr Trp Gln Trp Lys Asn Phe
185                      180                      185                      190
186
187 Ser Gly Leu Gln Glu Val Arg Ser His Val Pro Lys Phe Asp Ala Thr
188                      195                      200                      205
189
190 Trp Ala Thr Ala Arg Glu Val Thr Leu Lys Thr Phe Ala Glu Asp Asn
191      210                      215                      220
192
193 Ser Ala Ser Val Gln Ala Thr Met Tyr Lys Met Ala Glu Gln Ile Leu
194      225                      230                      235                      240
195
196 Ala Arg Gln Gln Leu Ile Glu Thr Val Glu Tyr Ser Leu Pro Asn Lys
197                      245                      250                      255
198
199 His Tyr Phe Glu Ile Asp Leu Ser Trp His Lys Gly Leu Gln Asn Thr
200                      260                      265                      270
201
202 Gly Lys Asn Ala Glu Val Phe Ala Pro Gln Ser Asp Pro Asn Gly Leu
203                      275                      280                      285
204
205 Ile Lys Cys Thr Val Gly Arg Ser Ser Leu Lys Ser Lys Leu
206      290                      295                      300
207
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 906 base pairs

(B) TYPE: nucleic acid

Patent Application US/07/659,408A

213 (C) STRANDEDNESS: single

214 (D) TOPOLOGY: linear

215

216 (ii) MOLECULE TYPE: DNA (genomic)

217

218

219 (vii) IMMEDIATE SOURCE:

220 (B) CLONE: Preferred sequence for expression in
221 prokaryotes
222

223

224

224 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

225

226 ATGTCTGCGG TAAAAGCAGC GCGCTACGGC AAGGACAATG TTCGCGTCTA CAAGGTTTCAC 60

227

228 AAGGACGAGA AGACCGGTGT CCAGACGGTG TACGAGATGA CCGTCTGTGT GCTTCTGGAG 120

229

230 GGTGAGATTG AGACCTCTTA CACCAAGGCC GACAACAGCG TCATTGTCGC AACCGACTCC 180

231

232 ATTAAGAACA CCATTTACAT CACCGCCAAG CAGAACCCCG TTACTCCTCC CGAGCTGTTC 240

233

234 GGCTCCATCC TGGGCACACA CTTCAATTGAG AAGTACAACC ACATCCATGC CGCTCACGTC 300

235

236 AACATTGTCT GCCACCGCTG GACCCGGATG GACATTGACG GCAAGCCACA CCCTCACTCC 360

237

238 TTCATCCGCG ACAGCGAGGA GAAGCGGAAT GTGCAGGTGG ACGTGGTCGA GGGCAAGGGC 420

239

240 ATCGATATCA AGTCGTCTCT GTCCGGCCTG ACCGTGCTGA AGAGCACCAA CTCGCAGTTC 480

241

242 TGGGGCTTCC TGCGTGACGA GTACACCACA CTTAAGGAGA CCTGGGACCG TATCCTGAGC 540

243

244 ACCGACGTCG ATGCCACTTG GCAGTGGAAG AATTTCAGTG GACTCCAGGA GGTCCGCTCG 600

245

246 CACGTGCCTA AGTTCGATGC TACCTGGGCC ACTGCTCGCG AGGTCACTCT GAAGACTTTT 660

247

248 GCTGAAGATA ACAGTGCCAG CGTGCAGGCC ACTATGTACA AGATGGCAGA GCAAATCCTG 720

249

250 GCGCGCCAGC AGCTGATCGA GACTGTGAG TACTCGTTGC CTAACAAGCA CTATTTGAA 780

251

252 ATCGACCTGA GCTGGCACAA GGGCCTCCAA AACACCGGCA AGAACGCCGA GGTCTTCGCT 840

253

254 CCTCAGTCGG ACCCCAACGG TCTGATCAAG TGTACCGTCG GCCGGTCCTC TCTGAAGTCT 900

255

256 AAATTG 906

257

258 (2) INFORMATION FOR SEQ ID NO:4:

259

260 (i) SEQUENCE CHARACTERISTICS:

261 (A) LENGTH: 906 base pairs

262 (B) TYPE: nucleic acid

263 (C) STRANDEDNESS: single

264 (D) TOPOLOGY: linear

265

Patent Application US/07/659,408A

266 (ii) MOLECULE TYPE: DNA (genomic)

267

268

269 (vii) IMMEDIATE SOURCE:

270 (B) CLONE: Preferred sequence for expression in
271 eukaryotes

272

273

274 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

275

276 ATGTCTGCTG TTAAGGCTGC TAGATACGGT AAGGACAACG TTAGAGTCTA CAAGGTTTAC 60

277

278 AAGGACGAGA AGACCGGTGT CCAGACGGTG TACGAGATGA CCGTCTGTGT GCTTCTGGAG 120

279

280 GGTGAGATTG AGACCTCTTA CACCAAGGCC GACAACAGCG TCATTGTCGC AACCGACTCC 180

281

282 ATTAAGAACA CCATTACAT CACCGCCAAG CAGAACCCCG TTACTCCTCC CGAGCTGTTC 240

283

284 GGCTCCATCC TGGGCACACA CTTATTGAG AAGTACAACC ACATCCATGC CGCTCACGTC 300

285

286 AACATTGTCT GCCACCGCTG GACCCGGATG GACATTGACG GCAAGCCACA CCCTCACTCC 360

287

288 TTCATCCGCG ACAGCGAGGA GAAGCGGAAT GTGCAGGTGG ACGTGGTCGA GGGCAAGGGC 420

289

290 ATCGATATCA AGTCGTCTCT GTCCGGCCTG ACCGTGCTGA AGAGCACCAA CTCGCAGTTC 480

291

292 TGGGGCTTCC TGCGTGACGA GTACACCACA CTTAAGGAGA CCTGGGACCG TATCCTGAGC 540

293

294 ACCGACGTCG ATGCCACTTG GCAGTGGAAG AATTTAGTG GACTCCAGGA GGTCCGCTCG 600

295

296 CACGTGCCTA AGTTCGATGC TACCTGGGCC ACTGCTCGCG AGGTCACTCT GAAGACTTTT 660

297

298 GCTGAAGATA ACAGTGCCAG CGTGCAGGCC ACTATGTACA AGATGGCAGA GCAAATCCTG 720

299

300 GCGCGCCAGC AGCTGATCGA GACTGTGCGAG TACTCGTTGC CTAACAAGCA CTATTTGAA 780

301

302 ATCGACCTGA GCTGGCACAA GGGCCTCCAA AACACCGGCA AGAACGCCGA GGTCTTCGCT 840

303

304 CCTCAGTCGG ACCCCAACGG TCTGATCAAG TGTACCGTCG GCCGGTCCTC TCTGAAGTCT 900

305

306 AAATTG 906

307

308 (2) INFORMATION FOR SEQ ID NO:5:

309

310 (i) SEQUENCE CHARACTERISTICS:

311 (A) LENGTH: 14 base pairs

312 (B) TYPE: nucleic acid

313 (C) STRANDEDNESS: single

314 (D) TOPOLOGY: linear

315

316 (ii) MOLECULE TYPE: DNA (genomic)

317

318 (iii) HYPOTHETICAL: NO

Patent Application US/07/659,408A

319

320

321 (vii) IMMEDIATE SOURCE:

322 (B) CLONE: Preferred non-translated 5' sequence for
323 animal cells

324

325

326 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

327

328 AGCTTGCCGC CACT

14

329

330 (2) INFORMATION FOR SEQ ID NO:6:

331

332 (i) SEQUENCE CHARACTERISTICS:

333 (A) LENGTH: 906 base pairs

334 (B) TYPE: nucleic acid

335 (C) STRANDEDNESS: double

336 (D) TOPOLOGY: linear

337

338 (ii) MOLECULE TYPE: DNA (genomic)

339

340 (iii) HYPOTHETICAL: NO

341

342

343 (vii) IMMEDIATE SOURCE:

344 (B) CLONE: Preferred sequence for expression in animal
345 cells

346

347

348 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

349

350 ATGTCCGCAG TAAAAGCAGC CCGCTACGGC AAGGACAATG TCCGCGTCTA CAAGGTTTAC 60

351

352 AAGGACGAGA AGACCGGTGT CCAGACGGTG TACGAGATGA CCGTCTGTGT GCTTCTGGAG 120

353

354 GGTGAGATTG AGACCTCTTA CACCAAGGCC GACAACAGCG TCATTGTCGC AACCGACTCC 180

355

356 ATTAAGAACA CCATTTACAT CACCGCCAAG CAGAACCCCG TTACTCCTCC CGAGCTGTTC 240

357

358 GGCTCCATCC TGGGCACACA CTTATTGAG AAGTACAACC ACATCCATGC CGCTCACGTC 300

359

360 AACATTGTCT GCCACCGCTG GACCCGGATG GACATTGACG GCAAGCCACA CCCTCACTCC 360

361

362 TTCATCCGCG ACAGCGAGGA GAAGCGGAAT GTGCAGGTGG ACGTGGTCGA GGGCAAGGGC 420

363

364 ATCGATATCA AGTCGTCTCT GTCCGGCCTG ACCGTGCTGA AGAGCACCAA CTCGCAGTTC 480

365

366 TGGGGCTTCC TCGTGACGA GTACACCACA CTTAAGGAGA CCTGGGACCG TATCCTGAGC 540

367

368 ACCGACGTCG ATGCCACTTG GCAGTGGAAG AATTTAGTG GACTCCAGGA GGTCCGCTCG 600

369

370 CACGTGCCTA AGTTCGATGC TACCTGGGCC ACTGCTCGCG AGGTCACTCT GAAGACTTTT 660

371

Patent Application US/07/659,408A

372 GCTGAAGATA ACAGTGCCAG CGTGCAGGCC ACTATGTACA AGATGGCAGA GCAAATCCTG 720
373
374 GCGCGCCAGC AGCTGATCGA GACTGTCGAG TACTCGTTGC CTAACAAGCA CTATTTTCGAA 780
375
376 ATCGACCTGA GCTGGCACAA GGGCCTCCAA AACACCGGCA AGAACGCCGA GGTCTTCGCT 840
377
378 CCTCAGTCGG ACCCCAACGG TCTGATCAAG TGTACCGTCG GCCGGTCCTC TCTGAAGTCT 900
379
380 AAATTG 906
381

382 (2) INFORMATION FOR SEQ ID NO:7:

383

384 (i) SEQUENCE CHARACTERISTICS:

385 (A) LENGTH: 23 base pairs

386 (B) TYPE: nucleic acid

387 (C) STRANDEDNESS: single

388 (D) TOPOLOGY: linear

389

390 (ii) MOLECULE TYPE: DNA (genomic)

391

392 (iii) HYPOTHETICAL: NO

393

394

395 (vii) IMMEDIATE SOURCE:

396 (B) CLONE: reverse transcription primer

397

398

399 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

400

401 GATCCGGGCC CTTTTTTTTT TTT

23

402

403 (2) INFORMATION FOR SEQ ID NO:8:

404

405 (i) SEQUENCE CHARACTERISTICS:

406 (A) LENGTH: 10 amino acids

407 (B) TYPE: amino acid

408 (C) STRANDEDNESS: single

409 (D) TOPOLOGY: linear

410

411 (ii) MOLECULE TYPE: peptide

412

413 (iii) HYPOTHETICAL: NO

414

415

416 (vii) IMMEDIATE SOURCE:

417 (B) CLONE: Hydrolysis product T 17

418

419

420 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

421

422 Asn Val Gln Val Asp Val Val Glu Gly Lys

423 1 5 10

424

Patent Application US/07/659,408A

425 (2) INFORMATION FOR SEQ ID NO:9:
426
427 (i) SEQUENCE CHARACTERISTICS:
428 (A) LENGTH: 8 amino acids
429 (B) TYPE: amino acid
430 (C) STRANDEDNESS: single
431 (D) TOPOLOGY: linear
432
433 (ii) MOLECULE TYPE: peptide
434
435 (iii) HYPOTHETICAL: NO
436
437
438 (vii) IMMEDIATE SOURCE:
439 (B) CLONE: Hydrolysis product T 20
440
441
442 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
443
444 Asn Phe Ser Gly Leu Gln Glu Val
445 1 5
446
447 (2) INFORMATION FOR SEQ ID NO:10:
448
449 (i) SEQUENCE CHARACTERISTICS:
450 (A) LENGTH: 6 amino acids
451 (B) TYPE: amino acid
452 (C) STRANDEDNESS: single
453 (D) TOPOLOGY: linear
454
455 (ii) MOLECULE TYPE: peptide
456
457 (iii) HYPOTHETICAL: NO
458
459
460 (vii) IMMEDIATE SOURCE:
461 (B) CLONE: Hydrolysis product T 23
462
463
464 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:
465
466 Phe Asp Ala Thr Trp Ala
467 1 5
468
469 (2) INFORMATION FOR SEQ ID NO:11:
470
471 (i) SEQUENCE CHARACTERISTICS:
472 (A) LENGTH: 8 amino acids
473 (B) TYPE: amino acid
474 (C) STRANDEDNESS: single
475 (D) TOPOLOGY: linear
476
477 (ii) MOLECULE TYPE: peptide

Patent Application US/07/659,408A

478
479 (iii) HYPOTHETICAL: NO
480
481
482 (vii) IMMEDIATE SOURCE:
483 (B) CLONE: Hydrolysis product T 27
484
485
486 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
487
488 His Tyr Phe Glu Ile Asp Leu Ser
489 1 5
490
491 (2) INFORMATION FOR SEQ ID NO:12:
492
493 (i) SEQUENCE CHARACTERISTICS:
494 (A) LENGTH: 13 amino acids
495 (B) TYPE: amino acid
496 (C) STRANDEDNESS: single
497 (D) TOPOLOGY: linear
498
499 (ii) MOLECULE TYPE: peptide
500
501 (iii) HYPOTHETICAL: NO
502
503
504 (vii) IMMEDIATE SOURCE:
505 (B) CLONE: Hydrolysis product T 28
506
507
508 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
509
510 Ile Leu Ser Thr Asp Val Asp Ala Thr Trp Gln Trp Lys
511 1 5 10
512
513 (2) INFORMATION FOR SEQ ID NO:13:
514
515 (i) SEQUENCE CHARACTERISTICS:
516 (A) LENGTH: 11 amino acids
517 (B) TYPE: amino acid
518 (C) STRANDEDNESS: single
519 (D) TOPOLOGY: linear
520
521 (ii) MOLECULE TYPE: peptide
522
523 (iii) HYPOTHETICAL: NO
524
525
526 (vii) IMMEDIATE SOURCE:
527 (B) CLONE: Hydrolysis product T 29
528
529
530 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Patent Application US/07/659,408A

531

532 His Tyr Phe Glu Ile Asp Leu Ser Trp His Lys

533 1 5 10

534

535 (2) INFORMATION FOR SEQ ID NO:14:

536

537 (i) SEQUENCE CHARACTERISTICS:

538 (A) LENGTH: 11 amino acids

539 (B) TYPE: amino acid

540 (C) STRANDEDNESS: single

541 (D) TOPOLOGY: linear

542

543 (ii) MOLECULE TYPE: peptide

544

545 (iii) HYPOTHETICAL: NO

546

547

548 (vii) IMMEDIATE SOURCE:

549 (B) CLONE: Hydrolysis product T 31

550

551

552 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

553

554 Ser Thr Asn Ser Gln Phe Trp Gly Phe Leu Arg

555 1 5 10

556

557 (2) INFORMATION FOR SEQ ID NO:15:

558

559 (i) SEQUENCE CHARACTERISTICS:

560 (A) LENGTH: 16 amino acids

561 (B) TYPE: amino acid

562 (C) STRANDEDNESS: single

563 (D) TOPOLOGY: linear

564

565 (ii) MOLECULE TYPE: peptide

566

567 (iii) HYPOTHETICAL: NO

568

569

570 (vii) IMMEDIATE SOURCE:

571 (B) CLONE: Hydrolysis product T 32

572

573

574 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

575

576 Gln Asn Pro Val Thr Pro Pro Glu Leu Phe Gly Ser Ile Leu Gly Thr

577 1 5 10 15

578

579

580 (2) INFORMATION FOR SEQ ID NO:16:

581

582 (i) SEQUENCE CHARACTERISTICS:

583 (A) LENGTH: 16 amino acids

Patent Application US/07/659,408A

584 (B) TYPE: amino acid
585 (C) STRANDEDNESS: single
586 (D) TOPOLOGY: linear
587
588 (ii) MOLECULE TYPE: peptide
589
590 (iii) HYPOTHETICAL: NO
591
592
593 (vii) IMMEDIATE SOURCE:
594 (B) CLONE: Hydrolysis product T 33
595
596
597 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:
598
599 Gln Asn Pro Val Thr Pro Pro Glu Leu Phe Gly Ser Ile Leu Gly Thr
600 1 5 10 15
601
602
603 (2) INFORMATION FOR SEQ ID NO:17:
604
605 (i) SEQUENCE CHARACTERISTICS:
606 (A) LENGTH: 17 amino acids
607 (B) TYPE: amino acid
608 (C) STRANDEDNESS: single
609 (D) TOPOLOGY: linear
610
611 (ii) MOLECULE TYPE: peptide
612
613 (iii) HYPOTHETICAL: NO
614
615
616 (vii) IMMEDIATE SOURCE:
617 (B) CLONE: Hydrolysis product V 1
618
619
620 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:
621
622 Tyr Ser Leu Pro Asn Lys His Tyr Phe Glu Ile Asp Leu Ser Trp His
623 1 5 10 15
624
625 Lys
626
627
628 (2) INFORMATION FOR SEQ ID NO:18:
629
630 (i) SEQUENCE CHARACTERISTICS:
631 (A) LENGTH: 16 amino acids
632 (B) TYPE: amino acid
633 (C) STRANDEDNESS: single
634 (D) TOPOLOGY: linear
635
636 (ii) MOLECULE TYPE: peptide

Patent Application US/07/659,408A

637
638 (iii) HYPOTHETICAL: NO
639
640
641 (vii) IMMEDIATE SOURCE:
642 (B) CLONE: Hydrolysis product V 2
643
644
645 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:
646
647 Val Thr Leu Lys Thr Phe Ala Glu Asp Asn Ser Ala Ser Val Gln Ala
648 1 5 10 15
649
650
651 (2) INFORMATION FOR SEQ ID NO:19:
652
653 (i) SEQUENCE CHARACTERISTICS:
654 (A) LENGTH: 24 amino acids
655 (B) TYPE: amino acid
656 (C) STRANDEDNESS: single
657 (D) TOPOLOGY: linear
658
659 (ii) MOLECULE TYPE: peptide
660
661 (iii) HYPOTHETICAL: NO
662
663
664 (vii) IMMEDIATE SOURCE:
665 (B) CLONE: Hydrolysis product V 3
666
667
668 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:
669
670 Thr Ser Tyr Thr Lys Ala Asp Asn Ser Val Ile Val Asp Thr Asp Ser
671 1 5 10 15
672
673 Ile Lys Asn Thr Ile Tyr Ile Thr
674 20
675
676 (2) INFORMATION FOR SEQ ID NO:20:
677
678 (i) SEQUENCE CHARACTERISTICS:
679 (A) LENGTH: 28 amino acids
680 (B) TYPE: amino acid
681 (C) STRANDEDNESS: single
682 (D) TOPOLOGY: linear
683
684 (ii) MOLECULE TYPE: peptide
685
686 (iii) HYPOTHETICAL: NO
687
688
689 (vii) IMMEDIATE SOURCE:

Patent Application US/07/659,408A

690 (B) CLONE: Hydrolysis product V 5

691

692

693 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

694

695 Gly Lys Gly Ile Asp Ile Lys Ser Ser Leu Ser Gly Leu Thr Val Leu

696 1 5 10 15

697

698 Lys Ser Thr Asn Ser Gln Phe Trp Gly Phe Leu Arg

699 20 25

700

701 (2) INFORMATION FOR SEQ ID NO:21:

702

703 (i) SEQUENCE CHARACTERISTICS:

704 (A) LENGTH: 17 amino acids

705 (B) TYPE: amino acid

706 (C) STRANDEDNESS: single

707 (D) TOPOLOGY: linear

708

709 (ii) MOLECULE TYPE: peptide

710

711 (iii) HYPOTHETICAL: NO

712

713

714 (vii) IMMEDIATE SOURCE:

715 (B) CLONE: Hydolysis product V 6

716

717

718 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

719

720 Gly Lys Gly Ile Asp Ile Lys Ser Ser Leu Ser Gly Leu Thr Val Leu

721 1 5 10 15

722

723 Lys

724

725

726 (2) INFORMATION FOR SEQ ID NO:22:

727

728 (i) SEQUENCE CHARACTERISTICS:

729 (A) LENGTH: 1236 base pairs

730 (B) TYPE: nucleic acid

731 (C) STRANDEDNESS: single

732 (D) TOPOLOGY: linear

733

734 (ii) MOLECULE TYPE: DNA (genomic)

735

736 (iii) HYPOTHETICAL: NO

737

738

739 (vii) IMMEDIATE SOURCE:

740 (B) CLONE: Fragment 3

741

742

Patent Application US/07/659,408A

743 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

744
745 GATCCGCGGA AGCATAAAGT GTAAAGCCTG GGGTGCCTAA TGAGTGAGCT AACTTACATT 60
746
747 AATTGCGTTG CGCTCACTGC CCGCTTTCCA GTCGGGAAAC CTGTCGTGCC AGCTGCATTA 120
748
749 ATGAATCGGC CAACGCGCGG GGAGAGGCGG TTTGCGTATT GGGCGCCAGG GTGGTTTTTC 180
750
751 TTTTCACCAG TGAGACGGGC AACAGCTGAT TGCCCTTCAC CGCCTGGCCC TGAGAGAGTT 240
752
753 GCAGCAAGCG GTCCACGCTG GTTTGCCCCA CCACCCGAAA ATCCTGTTTG ATGGTGGTTA 300
754
755 ACGGCGGGAT ATAACATGAG CTGTCTTCGG TATCGTCGTA TCCCACTACC GAGATATCCG 360
756
757 CACCAACGCG CAGCCCGGAC TCGGTAATGG CGCGCATTGC GCCCAGCGCC ATCTGATCGT 420
758
759 TGGCAACCAG CATCGCAGTG GGAACGATGC CCTCATTCAG CATTTGCATG GTTTGTTGAA 480
760
761 AACCGGACAT GGCCTCCAG TCGCCTTCCC GTTCCGCTAT CGGCTGAATT TGATTGCGAG 540
762
763 TGAGATATTT ATGCCAGCCA GCCAGACGCA GACGCGCCGA GACAGAACTT AATGGGCCCCG 600
764
765 CTAACAGCGC GATTTGCTGG TGACCCAATG CGACCAGATG CTCCACGCCC AGTCGCGTAC 660
766
767 CGTCTTCATG GGAGAAAATA AACTGTGGA TGGGTGTCTG GTCAGAGACA TCAAGAAATA 720
768
769 ACGCCGGAAC ATTAGTGAG GCAGCTTCCA CAGCAATGGC ATCCTGGTCA TCCAGCGGAT 780
770
771 AGTTAATGAT CAGCCCCTG ACGCGTTGCG CGAGAAGATT GTGCACCGCC GCTTTACAGG 840
772
773 CTTGACGCC GCTTCGTTCT ACCATCGACA CCACCACGCT GGCACCCAGT TGATCGGCGC 900
774
775 GAGATTTAAT CGCCGCGACA ATTTGCGACG GCGCGTGAG GGCAGACTG GAGGTGGCAA 960
776
777 CGCCAATCAG CAACGACTGT TTGCCCCGCA GTTGTGTGTC CACGCGGTTG GGAATGTAAT 1020
778
779 TCAGCTCCGC CATCGCCGCT TCCACTTTTT CCCGCGTTTT CGCAGAAACG TGGCTGGCCT 1080
780
781 GGTTCACCAC GCGGGAAACG GTCTGATAAC AGACACCGGC AACTCTGCG ACATCGTATA 1140
782
783 ACGTTACTGG TTTCACATTC ACCACCCTGA ATTGACTCTC TTCCGGGCGC TATCATGCCA 1200
784
785 TACCGCGAAA GGTTTTGCGC CATTCGATGG TGTCCG 1236
786

787 (2) INFORMATION FOR SEQ ID NO:23:

788

789 (i) SEQUENCE CHARACTERISTICS:

790 (A) LENGTH: 321 base pairs

791 (B) TYPE: nucleic acid

792 (C) STRANDEDNESS: double

793 (D) TOPOLOGY: linear

794

795 (ii) MOLECULE TYPE: DNA (genomic)

Patent Application US/07/659,408A

796
797 (iii) HYPOTHETICAL: NO
798
799
800 (vii) IMMEDIATE SOURCE:
801 (B) CLONE: Fragment 4
802
803 (ix) FEATURE:
804 (A) NAME/KEY: CDS
805 (B) LOCATION: 107..316
806 (D) OTHER INFORMATION: /product= "regulatory signal + aa
807 1-44 human growth hormone precursor"
808
809
810 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:
811
812 TCGAGCTGAC TGACCTGTTG CTTATATTAC ATCGATAGCG TATAATGTGT GGAATTGTGA 60
813
814 GCGATAACAA TTTCACACAG TTTAACTTTA AGAAGGAGAT ATACAT ATG GCT ACC 115
815 Met Ala Thr
816 1
817
818 GGA TCC CGG ACT AGT CTG CTC CTG GCT TTT GGC CTG CTC TGC CTG CCC 163
819 Gly Ser Arg Thr Ser Leu Leu Leu Ala Phe Gly Leu Leu Cys Leu Pro
820 5 10 15
821
822 TGG CTT CAA GAG GGC AGT GCC TTC CCA ACC ATT CCC TTA TCT AGA CTT 211
823 Trp Leu Gln Glu Gly Ser Ala Phe Pro Thr Ile Pro Leu Ser Arg Leu
824 20 25 30 35
825
826 TTT GAC AAC GCT ATG CTC CGC GCC CAT CGT CTG CAC CAG CTG GCC TTT 259
827 Phe Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe
828 40 45 50
829
830 GAC ACC TAC CAG GAG TTT GAA GAA GCC TAT ATC CCA AAG GAA CAG AAG 307
831 Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys
832 55 60 65
833
834 TAT TCA TTC CTGCA 321
835 Tyr Ser Phe
836 70
837
838
839 (2) INFORMATION FOR SEQ ID NO:24:
840
841 (i) SEQUENCE CHARACTERISTICS:
842 (A) LENGTH: 70 amino acids
843 (B) TYPE: amino acid
844 (D) TOPOLOGY: linear
845
846 (ii) MOLECULE TYPE: protein
847
848 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Patent Application US/07/659,408A

849
850 Met Ala Thr Gly Ser Arg Thr Ser Leu Leu Leu Ala Phe Gly Leu Leu
851 1 5 10 15
852
853 Cys Leu Pro Trp Leu Gln Glu Gly Ser Ala Phe Pro Thr Ile Pro Leu
854 20 25 30
855
856 Ser Arg Leu Phe Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln
857 35 40 45
858
859 Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys
860 50 55 60
861
862 Glu Gln Lys Tyr Ser Phe
863 65 70
864

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 74 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: ClaI-NdeI fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CGATAGCGTA TAATGTGTGG AATTGTGAGC GGATAACAAT TTCACACAGT TTTTCGCGAA 60

GAAGGAGATA TACA 74

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 190 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

Patent Application US/07/659,408A

902 (B) CLONE: Plasmid p373,2 fragment
903
904
905 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:
906
907 GATCTTCAAG CAGACCTACA GCAAGTTCGA CACAACTCA CACAACGATG ACGCACTACT 60
908
909 CAAGAACTAC GGGCTGCTCT ACTGCTTCAG GAAGGACATG GACAAGGTCG AGACATTCCT 120
910
911 GCGCATCGTG CAGTGCCGCT CTGTGGAGGG CAGCTGTGGC TTCTAGTAAG GTACCCTGCC 180
912
913 CTACGTACCA 190
914
915 (2) INFORMATION FOR SEQ ID NO:27:
916
917 (i) SEQUENCE CHARACTERISTICS:
918 (A) LENGTH: 48 base pairs
919 (B) TYPE: nucleic acid
920 (C) STRANDEDNESS: single
921 (D) TOPOLOGY: linear
922
923 (ii) MOLECULE TYPE: DNA (genomic)
924
925 (iii) HYPOTHETICAL: NO
926
927
928 (vii) IMMEDIATE SOURCE:
929 (B) CLONE: AccI-NdeI synthetic fragment
930
931
932 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:
933
934 TATGTCTGCG GTAAAAGCAG CGCGCTACGG CAAGGACAAT GTTCGCGT 48
935
936 (2) INFORMATION FOR SEQ ID NO:28:
937
938 (i) SEQUENCE CHARACTERISTICS:
939 (A) LENGTH: 360 base pairs
940 (B) TYPE: nucleic acid
941 (C) STRANDEDNESS: single
942 (D) TOPOLOGY: linear
943
944 (ii) MOLECULE TYPE: DNA (genomic)
945
946 (iii) HYPOTHETICAL: NO
947
948
949 (vii) IMMEDIATE SOURCE:
950 (B) CLONE: Plasmid pEMR469 fragment
951
952
953 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:
954

Patent Application US/07/659,408A

955 GGGACGCGTC TCCTCTGCCG GAACACCGGG CATCTCCAAC TTATAAGTTG GAGAAATAAG 60
956
957 AGAATTTTCAG ATTGAGAGAA TGAAAAAAAAA AAAAAAAAAA AAGGCAGAGG AGAGCATAGA 120
958
959 AATGGGGTTC ACTTTTGGT AAAGCTATAG CATGCCTATC ACATATAAAT AGAGTGCCAG 180
960
961 TAGCGACTTT TTTCACACTC GAGATACTCT TACTACTGCT CTCTTGTGT TTTTATCACT 240
962
963 TCTTGTTTCT TCTTGGTAAA TAGAATATCA AGCTACAAAA AGCATACAAT CAACTATCAA 300
964
965 CTATTAACTA TATCGATACC ATATGGATCC GTCGACTCTA GAGGATCGTC GACTCTAGAG 360
966
967
968 (2) INFORMATION FOR SEQ ID NO:29:
969
970 (i) SEQUENCE CHARACTERISTICS:
971 (A) LENGTH: 58 base pairs
972 (B) TYPE: nucleic acid
973 (C) STRANDEDNESS: double
974 (D) TOPOLOGY: linear
975
976 (ii) MOLECULE TYPE: DNA (genomic)
977
978 (iii) HYPOTHETICAL: NO
979
980
981 (vii) IMMEDIATE SOURCE:
982 (B) CLONE: Fragment C
983
984
985 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:
986
987 CGATATACAC AATGTCTGCT GTTAAGGCTG CTAGATACGG TAAGGACAAC GTTAGAGT 58
988
989 (2) INFORMATION FOR SEQ ID NO:30:
990
991 (i) SEQUENCE CHARACTERISTICS:
992 (A) LENGTH: 1013 base pairs
993 (B) TYPE: nucleic acid
994 (C) STRANDEDNESS: double
995 (D) TOPOLOGY: linear
996
997 (ii) MOLECULE TYPE: DNA (genomic)
998
999 (iii) HYPOTHETICAL: NO
1000
1001
1002 (vii) IMMEDIATE SOURCE:
1003 (B) CLONE: Fragment D
1004
1005
1006 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:
1007

Patent Application US/07/659,408A

1008 CTACAAGGTT CACAAGGACG AGAAGACCGG TGTCCAGACG GTGTACGAGA TGACCGTCTG 60
1009
1010 TGTGCTTCTG GAGGGTGAGA TTGAGACCTC TTACACCAAG GCCGACAACA GCGTCATTGT 120
1011
1012 CGCAACCGAC TCCATTAAGA ACACCATTTA CATCACC GCCG AAGCAGAACC CCGTTACTCC 180
1013
1014 TCCCGAGCTG TTCGGCTCCA TCCTGGGCAC ACACTTCATT GAGAAGTACA ACCACATCCA 240
1015
1016 TGCCGCTCAC GTCAACATTG TCTGCCACCG CTGGACCCGG ATGGACATTG ACGGCAAGCC 300
1017
1018 ACACCCTCAC TCCTTCATCC GCGACAGCGA GGAGAAGCGG AATGTGCAGG TGGACGTGGT 360
1019
1020 CGAGGGCAAG GGCATCGATA TCAAGTCGTC TCTGTCCGGC CTGACCGTGC TGAAGAGCAC 420
1021
1022 CAACTCGCAG TTCTGGGGCT TCCTGCGTGA CGAGTACACC AACTTAAGG AGACCTGGGA 480
1023
1024 CCGTATCCTG AGCACCGACG TCGATGCCAC TTGGCAGTGG AAGAATTTCA GTGGACTCCA 540
1025
1026 GGAGGTCCGC TCGCACGTGC CTAAGTTCGA TGCTACCTGG GCCACTGCTC GCGAGGTCAC 600
1027
1028 TCTGAAGACT TTTGCTGAAG ATAACAGTGC CAGCGTGCAG GCCACTATGT ACAAGATGGC 660
1029
1030 AGAGCAAATC CTGGCGCGCC AGCAGCTGAT CGAGACTGTC GAGTACTCGT TGCCTAACAA 720
1031
1032 GCACTATTTT GAAATCGACC TGAGCTGGCA CAAGGGCCTC CAAAACACCG GCAAGAACGC 780
1033
1034 CGAGGTCTTC GCTCCTCAGT CGGACCCCAA CGGTCTGATC AAGTGATCCG TCGGCCGGTC 840
1035
1036 CTCTCTGAAG TCTAAATTGT AAACCAACAT GATTCTCAGC TTCCGGAGTT TCCAAGGCAA 900
1037
1038 ACTGTATATA GTCTGGGATA GGGTATAGCA TTCATTCACT TGTTTTTTAC TTCCAAAAAA 960
1039
1040 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGC CCG 1013
1041

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 207 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: Synthetic GAL7 fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Patent Application US/07/659,408A

1061 CGCGTCTATA CTTCGGAGCA CTGTTGAGCG AAGGCTCATT AGATATATTT TCTGTCATTT 60
1062
1063 TCCTTAACCC AAAAATAAGG GAGAGGGTCC AAAAAGCGCT CGGACAACTG TTGACCGTGA 120
1064
1065 TCCGAAGGAC TGGCTATACA GTGTTACAA AATAGCCAAG CTGAAAATAA TGTGTAGCCT 180
1066
1067 TTAGCTATGT TCAGTTAGTT TGGCATG 207
1068
1069 (2) INFORMATION FOR SEQ ID NO:32:
1070
1071 (i) SEQUENCE CHARACTERISTICS:
1072 (A) LENGTH: 23 base pairs
1073 (B) TYPE: nucleic acid
1074 (C) STRANDEDNESS: single
1075 (D) TOPOLOGY: linear
1076
1077 (ii) MOLECULE TYPE: DNA (genomic)
1078
1079 (iii) HYPOTHETICAL: NO
1080
1081
1082 (vii) IMMEDIATE SOURCE:
1083 (B) CLONE: Modified XbaI-MluI adapter
1084
1085
1086 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:
1087
1088 CTAGGCTAGC GGGCCCGCAT GCA 23
1089
1090 (2) INFORMATION FOR SEQ ID NO:33:
1091
1092 (i) SEQUENCE CHARACTERISTICS:
1093 (A) LENGTH: 422 base pairs
1094 (B) TYPE: nucleic acid
1095 (C) STRANDEDNESS: single
1096 (D) TOPOLOGY: linear
1097
1098 (ii) MOLECULE TYPE: DNA (genomic)
1099
1100 (iii) HYPOTHETICAL: NO
1101
1102
1103 (vii) IMMEDIATE SOURCE:
1104 (B) CLONE: Plasmid pSE1 "site binding to HindIII"
1105 fragment
1106
1107
1108 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:
1109
1110 AGCTGGCTCG CATCTCTCCT TCACGCGCCC GCCGCCCTAC CTGAGGCCGC CATCCACGCC 60
1111
1112 GGTGAGTCGC GTTCTGCCGC CTCCCGCCTG TGGTGCCTCC TGAAGTGCCT CCGCCGTCTA 120
1113

Patent Application US/07/659,408A

1114 GGTAGGCTCC AAGGGAGCCG GACAAAGGCC CGGTCTCGAC CTGAGCTCTA AACTTACCTA 180
1115
1116 GACTCAGCCG GCTCTCCACG CTTTGCCTGA CCCTGCTTGC TCAACTCTAC GTCTTTGTTT 240
1117
1118 CGTTTTCTGT TCTGCGCCGT TACAACCTCA AGGTATGCGC TGGGACCTGG CAGGCGGCAT 300
1119
1120 CTGGGACCCC TAGGAAGGGC TTGGGGGTCC TCGTGCCCAA GGCAGGGAAC ATAGTGGTCC 360
1121
1122 CAGGAAGGGG AGCAGAGGCA TCAGGGTGTC CACTTTGTCT CCGCAGCTCC TGAGCCTGCA 420
1123
1124 GA 422
1125

1126 (2) INFORMATION FOR SEQ ID NO:34:

1127

1128 (i) SEQUENCE CHARACTERISTICS:

1129 (A) LENGTH: 77 base pairs

1130 (B) TYPE: nucleic acid

1131 (C) STRANDEDNESS: double

1132 (D) TOPOLOGY: linear

1133

1134 (ii) MOLECULE TYPE: DNA (genomic)

1135

1136 (iii) HYPOTHETICAL: NO

1137

1138

1139 (vii) IMMEDIATE SOURCE:

1140 (B) CLONE: Synthetic HindIII-"site binding to BamHI"
1141 fragment

1142

1143

1144 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

1145

1146 AGCTTGTCGA CTAATACGAC TCACTATAGG GCGGCCGCGG GCCCCTGCAG GAATTCGGAT 60

1147

1148 CCCCCGGGTG ACTGACT 77

1149

1150 (2) INFORMATION FOR SEQ ID NO:35:

1151

1152 (i) SEQUENCE CHARACTERISTICS:

1153 (A) LENGTH: 61 base pairs

1154 (B) TYPE: nucleic acid

1155 (C) STRANDEDNESS: double

1156 (D) TOPOLOGY: linear

1157

1158 (ii) MOLECULE TYPE: DNA (genomic)

1159

1160 (iii) HYPOTHETICAL: NO

1161

1162

1163 (vii) IMMEDIATE SOURCE:

1164 (B) CLONE: Synthetic HindIII-AccI fragment

1165

1166

Patent Application US/07/659,408A

1167 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

1168
1169 AGCTTGCCGC CACTATGTCC GCAGTAAAAG CAGCCCGCTA CGGCAAGGAC AATGTCCGCG 60

1170
1171 T 61

1172

1173 (2) INFORMATION FOR SEQ ID NO:36:

1174

1175 (i) SEQUENCE CHARACTERISTICS:

1176 (A) LENGTH: 920 base pairs

1177 (B) TYPE: nucleic acid

1178 (C) STRANDEDNESS: single

1179 (D) TOPOLOGY: linear

1180

1181 (ii) MOLECULE TYPE: DNA (genomic)

1182

1183 (iii) HYPOTHETICAL: NO

1184

1185

1186 (vii) IMMEDIATE SOURCE:

1187 (B) CLONE: HindIII-SnaBI fragment

1188

1189

1190 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

1191

1192 AGCTTGCCGC CACTATGTCC GCAGTAAAAG CAGCCCGCTA CGGCAAGGAC AATGTCCGCG 60

1193

1194 TCTACAAGGT TCACAAGGAC GAGAAGACCG GTGTCCAGAC GGTGTACGAG ATGACCGTCT 120

1195

1196 GTGTGCTTCT GGAGGGTGAG ATTGAGACCT CTTACACCAA GGCCGACAAC AGCGTCATTG 180

1197

1198 TCGCAACCGA CTCCATTAAG AACACCATT ACATCACCGC CAAGCAGAAC CCCGTTACTC 240

1199

1200 CTCCCGAGCT GTTCGGCTCC ATCCTGGGCA CACACTTCAT TGAGAAGTAC AACCACATCC 300

1201

1202 ATGCCGCTCA CGTCAACATT GTCTGCCACC GCTGGACCCG GATGGACATT GACGGCAAGC 360

1203

1204 CACACCCTCA CTCCTTCATC CGCGACAGCG AGGAGAAGCG GAATGTGCAG GTGGACGTGG 420

1205

1206 TCGAGGGCAA GGGCATCGAT ATCAAGTCGT CTCTGTCCGG CCTGACCGTG CTGAAGAGCA 480

1207

1208 CCAACTCGCA GTTCTGGGGC TTCCTGCGTG ACGAGTACAC CACACTTAAG GAGACCTGGG 540

1209

1210 ACCGTATCCT GAGCACCGAC GTCGATGCCA CTTGGCAGTG GAAGAATTTC AGTGGACTCC 600

1211

1212 AGGAGGTCCG CTCGCACGTG CCTAAGTTCG ATGCTACCTG GGCCACTGCT CGCGAGGTCA 660

1213

1214 CTCTGAAGAC TTTTGCTGAA GATAACAGTG CCAGCGTGCA GGCCACTATG TACAAGATGG 720

1215

1216 CAGAGCAAAT CCTGGCGCGC CAGCAGCTGA TCGAGACTGT CGAGTACTCG TTGCCTAACA 780

1217

1218 AGCACTATTT CGAAATCGAC CTGAGCTGGC ACAAGGGCCT CCAAAACACC GGCAAGAACG 840

1219

Patent Application US/07/659,408A

1220	CCGAGGTCTT CGCTCCTCAG TCGGACCCCA ACGGTCTGAT CAAGTGTACC GTCGGCCGGT	900
1221		
1222	CCTCTCTGAA GTCTAAATTG	920

PAGE: 1

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION US/07/659,408A

DATE: 04/15/92
TIME: 09:07:56

LINE ERROR

ORIGINAL TEXT

38 Wrong application Serial Number
221 Response Exceeds Line Limitations
271 Response Exceeds Line Limitations
323 Response Exceeds Line Limitations
345 Response Exceeds Line Limitations
1105 Response Exceeds Line Limitations
1141 Response Exceeds Line Limitations

(A) APPLICATION NUMBER: US 07/659,408
prokaryotes
eukaryotes
animal cells
cells
fragment
fragment



PAGE: 1

SEQUENCE MISSING ITEM REPORT
PATENT APPLICATION US/07/659,408A

DATE: 04/15/92
TIME: 09:07:56

MANDATORY IDENTIFIER THAT WAS NOT FOUND

PRIOR APPLICATION DATA
APPLICATION NUMBER
FILING DATE

PAGE: 1

SEQUENCE CORRECTION REPORT
PATENT APPLICATION US/07/659,408A

DATE: 04/15/92
TIME: 09:07:56

LINE ORIGINAL TEXT

CORRECTED TEXT